WHAT IS CLAIMED IS:

1. A method for predicting at least one property of a candidate molecule, said method comprising:

classifying a set of reference molecules as either possessing or not possessing the at least one property;

selecting a subset of said set of reference molecules, wherein all of the molecules in said subset possess the at least one property;

selecting a plurality of marker molecules from the subset, said plurality of marker molecules being less in number than the number of molecules in said subset; and

comparing structural characteristics of said candidate molecule with structural characteristics of at least one of said marker molecules.

2. The method of Claim 1, wherein the at least one property is high protein binding.

3. A method of selecting a set of marker molecules for structural comparisons in a model for molecular behavior prediction, said method comprising:

classifying a set of reference molecules as either possessing or not possessing the at least one property;

selecting a subset of said set of reference molecules, wherein all of the molecules in said subset possess the at least one property;

selecting a plurality of marker molecules from the subset, said plurality of marker molecules being less in number than the number of molecules in said subset.

4. The method of Claim 3, wherein said subset comprises all of the molecules in said set that possess said at least one property.

5. The method of Claim 3, wherein said selecting a plurality of marker molecules comprises:

comparing all molecules in said set with all other molecules in said set in accordance with a pre-defined numerical similarity metric;

selecting a first molecule of said subset;

sorting all other molecules of said set in descending order of numerical similarity to said first molecule, thereby defining a similarity distance in terms of number of molecules between said first molecule and each other molecule of the set;

defining, for each range in molecules of similarity distance away from said first molecule, a fractions-correctly-predicted metric as the number of molecules in said range which are also members of said subset divided by the total number of molecules in said range;

counting the number of molecules away from said first molecule at which the fractions-correctly-predicted for said first molecule drops below a threshold value;

repeating said selecting, sorting, defining, and counting steps for all other molecules of said subset;

choosing, as said set of marker molecules, those molecules of said subset having a fractions-correctly-predicted metric which exceeds said threshold value for a pre-selected minimum distance.

6. The method of Claim 5, additionally comprising repeating said counting step for a plurality of different threshold values.

7. The method of Claim 6, comprising repeating said choosing step at a plurality of different threshold values and minimum distances so as to select a plurality of preliminary sets of marker molecules.

8. The method of Claim 7, comprising choosing a final set of marker molecules by making molecular behavior predictions for all molecules in said set using each one of said preliminary sets of marker molecules, and choosing as said final set of marker molecules the preliminary set that most accurately predicts molecular behavior of molecules of said set.

9. A method of predicting whether or not a molecule will be highly protein bound in serum, said method comprising:

numerically defining the structural similarity of said molecule to a plurality of marker molecules, all of which are known to be highly protein bound in serum;

comparing said structural similarities to a corresponding plurality of numerical thresholds associated with each of said plurality of marker molecules.

10. The method of Claim 9, wherein said molecule is categorized as highly protein bound if any one of the numerically defined structural similarities exceeds any corresponding one of said numerical thresholds.

11. The method of Claim 9, additionally comprising comparing the logP of said molecule to a logP threshold.

12. The method of Claim 11, wherein said molecule is categorized as highly protein bound if (1) any one of the numerically defined structural similarities exceeds any corresponding one of said numerical thresholds, or (2) if the logP of said molecule exceeds said logP threshold.

13. A system for predicting molecular activity, said system comprising:

one or more memories having stored thereon (1) structural information for a plurality of marker molecules, all of which possess a selected biological or chemical activity, (2) a numerical similarity threshold assigned to each of said plurality of marker molecules, and (3) structural information for at least one candidate molecule;

a processor configured to (1) structurally compare said at least one candidate molecule to all of said plurality of marker molecules to produce a set of numerical similarity metrics, and (2) compare said numerical similarity metrics with said numerical similarity thresholds.

14. The system of Claim 13, wherein said selected biological or chemical activity comprises high protein binding.

15. A system for predicting propensity for protein binding in a candidate molecule, said system comprising:

one or more memories storing structural information related to a plurality of marker molecules, all of which are known to be highly protein bound, and storing structural information related to said candidate molecule;

means for numerically defining the structural similarity of said candidate molecule to said plurality of marker molecules; and

means for comparing said structural similarities to a corresponding plurality of numerical thresholds associated with each of said plurality of marker molecules.

16. The system of Claim 15, additionally comprising means for estimating the logP of said candidate molecule and comparing said estimated logP to a threshold.